

Karing, Constance; Pfost, Maximilian; Artelt, Cordula

Hängt die diagnostische Kompetenz von Sekundarstufenlehrkräften mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz ihrer Schülerinnen und Schüler zusammen?

Journal for educational research online 3 (2011) 2, S. 119-147



Quellenangabe/ Reference:

Karing, Constance; Pfost, Maximilian; Artelt, Cordula: Hängt die diagnostische Kompetenz von Sekundarstufenlehrkräften mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz ihrer Schülerinnen und Schüler zusammen? - In: Journal for educational research online 3 (2011) 2, S. 119-147 - URN: urn:nbn:de:0111-opus-56261 - DOI: 10.25656/01:5626

<https://nbn-resolving.org/urn:nbn:de:0111-opus-56261>

<https://doi.org/10.25656/01:5626>

in Kooperation mit / in cooperation with:



WAXMANN
www.waxmann.com

<http://www.waxmann.com>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Constance Karing, Maximilian Pfost & Cordula Artelt

Hängt die diagnostische Kompetenz von Sekundarstufenlehrkräften mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz ihrer Schülerinnen und Schüler zusammen?

Zusammenfassung

In der vorliegenden Arbeit wird untersucht, ob die diagnostische Kompetenz von Sekundarstufenlehrkräften mit der Leistungsentwicklung von Schülerinnen und Schülern im Lesen und im Rechnen zusammenhängt und inwieweit dies durch Unterrichtsvariablen moderiert wird. Datenbasis ist eine Stichprobe von 502 Schülerinnen und Schülern sowie deren 40 Deutsch- und 29 Mathematiklehrkräfte, die zu zwei Messzeitpunkten (am Ende der 5. und 6. Jahrgangsstufe) im Rahmen der BiKS-Studie untersucht wurden. Der aufgabenspezifische Treffer und die Rangordnungskomponente werden als Indikatoren für die diagnostische Kompetenz der Lehrkräfte herangezogen. Die Ergebnisse zeigen, dass die Deutsch- und Mathematiklehrkräfte eine mittelhohe Trefferquote aufweisen (aufgabenspezifische Treffer). Für die Rangordnungskomponente fallen die durchschnittlichen Korrelationen mit $\bar{r} = .19$ für den Bereich Lesen und $\bar{r} = .44$ für den Bereich Rechnen vergleichsweise niedrig aus. Darüber hinaus ergeben die durchgeführten Mehrebenenanalysen einen signifikanten positiven Zusammenhang des aufgabenspezifischen Treffers mit der Entwicklung der Lesekompetenz, aber nicht mit der Entwicklung der mathematischen Kompetenz der Schülerinnen und Schüler. Allerdings zeigt sich, dass der Zusammenhang zwischen dem aufgabenspezifischen Treffer und der Lesekompetenzentwicklung durch Unterrichtsvariablen (Individualisierung, Strukturierungshilfen) moderiert wird. Für die Rangordnungskomponente kann-

Dipl.-Psych. Constance Karing (corresponding author), Friedrich-Schiller-Universität Jena, Institut für Psychologie, Abteilung für Forschungssynthese, Intervention und Evaluation, Humboldtstraße 26, 07743 Jena, Deutschland
E-Mail: constance.karing@uni-jena.de

Dr. Maximilian Pfost · Prof. Dr. Cordula Artelt, Otto-Friedrich-Universität Bamberg, Fakultät Humanwissenschaften, Lehrstuhl für Empirische Bildungsforschung, Markusplatz 3, 96045 Bamberg, Deutschland
E-Mail: maximilian.pfost@uni-bamberg.de
cordula.artelt@uni-bamberg.de

ten keine signifikanten positiven Zusammenhänge bzw. Wechselwirkungen für die beiden Bereiche nachgewiesen werden.

Schlagworte

Diagnostische Kompetenz; Kompetenzentwicklung; Unterrichtsqualität

Is there a relationship between lower secondary school teacher judgment accuracy and the development of students' reading and mathematical competence?

Abstract

The present study investigates the relationship between teacher judgment accuracy and the development of students' reading and mathematical competence and whether this relationship is moderated by instructional variables. Longitudinal data were obtained from a sample of 502 students and their 40 German language and 29 mathematics teachers in the context of the BiKS study (measurement points: at the end of grades 5 and 6). Teacher judgment accuracy was measured by a task-specific hit rate and the rank order component in the domains of reading and mathematics. German language and mathematics teachers showed a moderate hit rate. For the rank order component, we found a mean correlation of $\bar{r} = .19$ in the domain of reading and $\bar{r} = .44$ in the domain of mathematics. Multilevel analyses revealed that the task-specific hit rate had a significant positive relationship with the development of students' reading competence, but it was not related to the development of students' mathematical competence. Furthermore, the significant relationship of teachers' task-specific hit rate was moderated by instructional variables such as teachers' use of individualization and structuring cues. For the rank order component, no significant positive relationships or interactions were found in the domains of reading and mathematics.

Keywords

Teacher judgment accuracy; Competence development; Instructional quality

1. Einleitung

Die diagnostische Kompetenz von Lehrkräften wird als eine der professionellen Kompetenzen von Lehrkräften beschrieben (vgl. Baumert & Kunter, 2006; Ständiges Sekretariat der Kultusminister der Länder der Bundesrepublik Deutschland (KMK), 2004) und stellt nach Weinert, Schrader und Helmke (1990) neben der Klassenführungscompetenz, der fachwissenschaftlichen und fachdidaktischen Kompetenz eine der vier Schlüsselkompetenzen im Unterricht dar. Diagnostische Kompetenz bezeichnet dabei die Fähigkeit der Lehrkräfte, Schülermerkmale, aber auch Lern- und Aufgabenanforderungen zutreffend einzuschät-

zen, und wird daher mehr oder weniger mit der Diagnose- oder Urteilsgenauigkeit gleichgesetzt (vgl. Artelt & Gräsel, 2009; Helmke, Hosenfeld & Schrader, 2004). In einem weiteren Sinne ist damit auch die Fähigkeit gemeint, den Kenntnisstand, die Verarbeitungs- und Verstehensprozesse sowie evtl. Leistungsschwierigkeiten von Schülerinnen und Schülern zutreffend einzuschätzen, was wiederum eine wichtige Voraussetzung für eine optimale und individualisierte Förderung der Schülerinnen und Schüler darstellt (vgl. Artelt, Stanat, Schneider & Schiefele, 2001; Helmke, Hosenfeld & Schrader, 2003). Der diagnostischen Kompetenz kommt demnach ein zentraler Stellenwert bei der Unterrichtsgestaltung, aber ebenso bei pädagogischen Entscheidungen und Handlungen (z.B. Übertrittsempfehlung, Aufgabenwahl, Bewertung) zu (vgl. Artelt & Gräsel, 2009; Helmke et al., 2004).

1.1 Komponenten der Urteilsgenauigkeit

In der Forschung zur diagnostischen Kompetenz von Lehrkräften wird die Akkuratheit von Lehrerurteilen in der Regel über einen Vergleich des Lehrerurteils zu bestimmten Schülermerkmalen mit den tatsächlichen Merkmalsausprägungen der Schülerinnen und Schüler bestimmt. Für die Ermittlung dieser Übereinstimmungen werden in der Forschungsliteratur hauptsächlich drei Urteilskomponenten, die Rangordnungs-, die Differenzierungs- und die Niveauelemente, verwendet, die jeweils unterschiedliche Facetten der Urteilsgenauigkeit abbilden (vgl. Karing, Matthäi & Artelt, 2011; Schrader & Helmke, 1987; Spinath, 2005; Südkamp, Möller & Pohlmann, 2008). Nach Helmke und Kollegen (2004) bildet dabei die Rangordnungskomponente das Kernstück der diagnostischen Kompetenz ab. Sie beschreibt die Fähigkeit der Lehrkräfte, die Rangordnung bzw. Fähigkeitsabstufung zwischen verschiedenen Schülerinnen und Schülern zutreffend einzuschätzen. Dafür werden klassenspezifische Korrelationen zwischen den Lehrereinschätzungen und den tatsächlichen Schülerleistungen berechnet (vgl. Helmke et al., 2004). Die Mehrheit der Studien deutet darauf hin, dass die Lehrkräfte im Großen und Ganzen angemessene Einschätzungen über Schülerleistungen vornehmen (z.B. Hoge & Coladarci, 1989: Median von $r = .66$; Südkamp, Kaiser & Möller, 2011: Median von $r = .53$). Jedoch zeigen sich in diesen Studien auch erhebliche Unterschiede zwischen den einzelnen Lehrkräften hinsichtlich ihrer Diagnosegenauigkeit. Eine Ergänzung zu den klassischen Urteilskomponenten stellt der aufgabenspezifische Treffer dar, der einen exakten Abgleich von Lehrereinschätzung und Schülerleistung auf der Ebene einzelner Aufgaben beinhaltet (vgl. Karing et al., 2011). Die Besonderheit des aufgabenspezifischen Treffers liegt entsprechend darin, dass dabei beachtet wird, ob die von der Lehrkraft als gelöst (bzw. nicht gelöst) angegebenen Aufgaben auch tatsächlich dieselben Aufgaben sind, die die Schülerin bzw. der Schüler löste (bzw. nicht löste). Damit ist es außerdem möglich über diesen Indikator der diagnostischen Kompetenz festzustellen, ob Lehrkräfte die Schülerleistung je nach Schwierigkeit, Art der Anforderung oder Typ der Aufgabe unterschiedlich akku-

rat einschätzen können, was besonders für die Konzeption von Interventionen relevant ist (vgl. Karing et al., 2011). Grundlage für dieses Treffermaß sind aufgabenspezifische Lehrerurteile. Hierbei müssen Lehrkräfte sowohl die Schülerfähigkeit als auch die Schwierigkeit der entsprechenden Aufgabe zueinander in Beziehung setzen. Damit erfordern solche Einschätzungen sowohl Wissen über Personen als auch Wissen über Aufgabenmerkmale. Es ergibt sich somit eine Überlappung von diagnostischer, didaktischer und fachwissenschaftlicher Kompetenz (vgl. Helmke et al., 2003; Karing et al., 2011). Da Lehrkräfte die Schülerfähigkeiten mit den konkreten Aufgabenanforderungen abgleichen müssen, ist der kognitive Verarbeitungsaufwand bei derartigen Einschätzungen auch höher als bei globalen Fähigkeitseinschätzungen ohne Bezug zu konkreten Einzelanforderungen. Bisher liegen nur wenige Studien vor, die dieses Treffermaß betrachteten. In der COACTIV-Studie (vgl. Brunner, Anders, Hachfeld & Krauss, 2011) lag bei den Mathematiklehrkräften der Sekundarstufe die mittlere Trefferquote (für insgesamt zwei Aufgaben) bei 51 %. Studien bei Grundschullehrkräften im englischsprachigen Raum ergaben eine etwas höhere Übereinstimmung zwischen Lehrerurteil und Schülerleistung (vgl. Coladarci, 1986; Demaray & Elliott, 1998). So schätzten die Grundschullehrkräfte in der Studie von Coladarci (1986) sowohl im Bereich Lesekompetenz als auch in Mathematik im Mittel 73 % der Schülerantworten richtig ein. Demaray und Elliott (1998) fanden bei den Grundschullehrkräften im Bereich Lesekompetenz sowie in Mathematik sogar eine mittlere Trefferquote von 79 %. Allerdings zeigten sich in allen Studien beträchtliche Unterschiede zwischen den einzelnen Lehrkräften.

1.2 Relevanz der diagnostischen Kompetenz für den Lernerfolg der Schülerinnen und Schüler

Basierend unter anderem auf den Prozess-Mediations-Produkt-Modellen des Unterrichts wird in der Unterrichtsforschung angenommen, dass die diagnostische Kompetenz der Lehrkräfte relevant für den Lernfortschritt der Schülerinnen und Schüler ist (vgl. Brunner et al., 2011; Kuntze, 2006). Denn ausreichendes diagnostisches Wissen der Lehrkräfte ist erforderlich – so die Annahme – um den Unterricht an die Lernvoraussetzungen der Schülerinnen und Schüler durch differenzierte und individualisierte Maßnahmen anzupassen (vgl. Haag & Lohrmann, 2007; Schrader & Helmke, 2001). Dafür muss die Lehrkraft zum einen die lern- und leistungsrelevanten Schülermerkmale akkurat einschätzen und zum anderen auch das Schwierigkeitspotenzial des Unterrichtsmaterials (vgl. Anders, Kunter, Brunner, Krauss & Baumert, 2010) sowie die Schwierigkeit der eingesetzten Instruktionsstrategien und -methoden angemessen beurteilen können. Insgesamt ist somit für einen adaptiven sowie auch effektiven Unterricht eine enge Verknüpfung von diagnostischen Informationen und didaktischen Maßnahmen notwendig (vgl. Ingenkamp, 1992; Schrader & Helmke, 2001).

Bislang gibt es nur wenige empirische Arbeiten, die den Zusammenhang zwischen der diagnostischen Kompetenz und dem Lernerfolg der Schülerinnen und Schüler untersucht haben. Diese Studien beschränkten sich auf den mathematischen Bereich und kamen zu heterogenen Befunden. In der Studie von Schrader (1989; vgl. auch Schrader & Helmke, 1987) hing die diagnostische Kompetenz der Hauptschullehrkräfte nicht mit der Leistungsentwicklung der Schülerinnen und Schüler im Fach Mathematik in der fünften Jahrgangsstufe zusammen. Weder die personenbezogene noch die aufgabenbezogene Diagnosekompetenz (jeweils die Rangordnungskomponente) trugen zur Erklärung des Lernerfolgs der Schülerinnen und Schüler bei. Basis für die personenbezogene Diagnosekompetenz waren hierbei die Einschätzungen der Lehrkräfte, wie viele der Aufgaben des Mathematiktests jede einzelne Schülerin bzw. jeder einzelne Schüler lösen würde, während die aufgabenbezogene Diagnosekompetenz sich auf die Fähigkeit der Lehrkräfte bezog, die Schwierigkeit der jeweiligen Mathematikaufgabe in der Klasse zu beurteilen. Allerdings zeigte sich in dieser Studie eine signifikante Wechselwirkung zwischen den beiden Facetten der diagnostischen Kompetenz und den Strukturierungshilfen bzw. den individuellen fachlichen Zuwendungen. Strukturierungshilfen umfassten hierbei explizite Hinweise der Lehrkräfte zur Struktur des Unterrichtsstoffes und zur Relevanz bestimmter Informationen, während individuelle fachliche Zuwendungen sich auf die fachliche Unterstützung einzelner Schülerinnen und Schüler im Rahmen von Stillarbeitsphasen im Unterricht bezogen. Der Lernerfolg der Schülerinnen und Schüler war dann am größten, wenn eine hohe diagnostische Kompetenz mit einer hohen Ausprägung des jeweiligen Unterrichtsmerkmals einherging. Damit kam der diagnostischen Kompetenz der Lehrkräfte erst im Zusammenwirken mit geeigneten didaktischen Maßnahmen eine Bedeutung für den Lernerfolg zu. Dagegen gab es in der Arbeit von Lehmann und Kollegen (2000) Hinweise darauf, dass zumindest bei einzelnen Klassenstufen und einzelnen Schulformen die Fähigkeit der Lehrkräfte zur Beurteilung von Aufgabenschwierigkeiten in Bezug auf die eigene Klasse (Rangordnungskomponente) positiv mit der einmalig erhobenen Mathematikleistung der Schülerinnen und Schüler zusammenhing. Auch Anders und Kollegen (2010; Brunner et al., 2011) stellten sowohl für die Rangordnungskomponente als auch für den aufgabenbezogenen Urteilsfehler (mittlerer Absolutbetrag zwischen tatsächlichem Anteil richtiger Lösungen in der Klasse und den Lehrereinschätzungen) positive Zusammenhänge mit der Leistungsentwicklung der Schülerinnen und Schüler in Mathematik fest. So wiesen die Schülerinnen und Schüler bei gleichen Eingangsvoraussetzungen (u. a. allgemeine kognitive Fähigkeiten, sozioökonomischer Status) eine höhere Mathematikleistung in der zehnten Jahrgangsstufe auf, deren Lehrkräfte akkurater die Leistungsrangfolge der Schülerinnen und Schüler (Rangordnungskomponente) sowie die aufgabenbezogene Schwierigkeit (aufgabenbezogener Urteilsfehler) einschätzen konnten. Darüber hinaus stellten die Autoren fest, dass der direkte Zusammenhang zwischen der Fähigkeit, die Aufgabenschwierigkeit akkurat einschätzen zu können, und dem Lernerfolg der Schülerinnen und Schüler

teilweise durch ein höheres kognitives Aktivierungspotenzial im Unterricht erklärt werden konnte. Dagegen ließ sich der direkte Zusammenhang zwischen der Rangordnungskomponente und der Leistungsentwicklung nicht durch das Ausmaß der kognitiven Aktivierung im Unterricht erklären.

2. Forschungsfragen

Die bisherige Forschungsübersicht zeigt, dass nur wenige Studien den aufgabenspezifischen Treffer betrachten. Diese Arbeiten sind vor allem im Grundschulbereich angesiedelt. Darüber hinaus gibt es bislang kaum empirische Erkenntnisse über den Zusammenhang zwischen der diagnostischen Kompetenz und dem Lernerfolg der Schülerinnen und Schüler. Die wenigen Studien beschränken sich hierbei auf den mathematischen Bereich. Deshalb gehen wir im Rahmen dieses Beitrags folgenden Fragestellungen nach:

1. Wie akkurat schätzen die Deutsch- und Mathematiklehrkräfte die Schülerleistungen ein? Im Einklang mit der Forschungsliteratur (z.B. Coladarsi, 1986; Demaray & Elliott, 1998; Hoge & Coladarsi, 1989) erwarten wir sowohl im Bereich Lesen als auch im Bereich Rechnen eine moderate bis hohe Urteilsgüte.

2. Hängen die beiden Komponenten der diagnostischen Kompetenz mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz der Schülerinnen und Schüler zusammen? Basierend unter anderem auf den Prozess-Mediations-Produkt-Modellen der Unterrichtsforschung (z.B. Brunner et al., 2011; Kuntze, 2006) nehmen wir an, dass die Rangordnungskomponente und der aufgabenspezifische Treffer als Indikatoren der diagnostischen Kompetenz positiv mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz der Schülerinnen und Schüler zusammenhängen.

3. Werden die Zusammenhänge dieser beiden Urteilskomponenten mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz der Schülerinnen und Schüler durch die Unterrichtsvariablen Individualisierung und Strukturierungshilfen moderiert? In der Forschungsliteratur wird angenommen, dass die für effektiven Unterricht notwendige Anpassung des Unterrichts an die Lernvoraussetzungen der Schülerinnen und Schüler eine enge Verknüpfung von diagnostischem Wissen und didaktischen Maßnahmen voraussetzt. Hierbei wird davon ausgegangen, dass u. a. strukturierende sowie individualisierende Maßnahmen in Interaktion mit hoher diagnostischer Kompetenz eine lernförderliche Wirkung besitzen (vgl. Schrader, 1989). Dies konnte auch in der Arbeit von Schrader (1989; vgl. auch Schrader & Helmke, 1987) nachgewiesen werden. Der größte Lernerfolg der Schülerinnen und Schüler in Mathematik zeigte sich, wenn eine hohe diagnostische Kompetenz der Lehrkraft mit einer hohen Ausprägung des jeweiligen Unterrichtsmerkmals einherging. Vor diesem Hintergrund erwarten wir, dass der Zusammenhang zwischen den beiden Indikatoren der diagnostischen Kompetenz und der Entwicklung der Lesekompetenz sowie der mathematischen Kompetenz

der Schülerinnen und Schüler durch die Unterrichtsmerkmale Individualisierung und Strukturierungshilfen moderiert wird.

3. Methode

3.1 Stichprobe

Die Datengrundlage für die vorliegende Untersuchung bildet der Längsschnitt 8-14 der Bamberger DFG-Forschergruppe BiKS (Bildungsprozesse, Kompetenzentwicklung und Formation von Selektionsentscheidungen im Vor- und Grundschulalter), in dem Schülerinnen und Schüler bayerischer und hessischer Schulen in ihrer Entwicklung vom achten bis zum vierzehnten Lebensjahr begleitet werden. Die Schülerinnen und Schüler werden hierbei in regelmäßigen Abständen in verschiedenen schulbezogenen Bereichen getestet und umfassend befragt. Die vorliegende Arbeit bezieht sich auf Daten aus zwei Erhebungswellen zu Beginn der Sekundarstufe I, die im Schuljahr 2007/2008 am Ende der fünften Jahrgangsstufe (t_1) und im Schuljahr 2008/2009 am Ende der sechsten Jahrgangsstufe (t_2) durchgeführt wurden. Die Analysestichprobe umfasste 502 Schülerinnen und Schüler sowie deren 40 Deutschlehrkräfte und 29 Mathematiklehrkräfte aus 29 Schulen der Sekundarstufe (4 Hauptschulen, 5 Realschulen und 20 Gymnasien) in Bayern und Hessen (28 Schulklassen in Bayern, 12 Schulklassen in Hessen). In den durchgeführten Analysen wurden nur diejenigen Klassen berücksichtigt, die zu beiden Messzeitpunkten dieselbe Deutschlehrkraft bzw. dieselbe Mathematiklehrkraft hatten. Sowohl die Deutsch- als auch die Mathematiklehrkräfte waren zu t_1 im Durchschnitt 46.5 Jahre alt (Deutschlehrkräfte: $SD = 12.0$, Mathematiklehrkräfte: $SD = 10.3$). 57.5 % der Deutschlehrkräfte und 41.4 % der Mathematiklehrkräfte waren weiblich. Die Berufserfahrung der Deutschlehrkräfte lag bei durchschnittlich 17.8 Jahren ($SD = 11.1$) und die der Mathematiklehrkräfte bei 15.5 Jahren ($SD = 12.3$).

Die Stichprobe der Schüler umfasste 294 Mädchen und 208 Jungen. Das Durchschnittsalter zu t_1 lag bei 11.4 Jahren ($SD = 0.4$). Von den Schülerinnen und Schülern besuchten 50 (10.0 %) die Hauptschule, 63 (12.5 %) die Realschule und 389 (77.5 %) das Gymnasium.

3.2 Instrumente

3.2.1 Schülervariablen

3.2.1.1 Lesekompetenz. Die Lesekompetenz der Schülerinnen und Schüler wurde am Ende der fünften Jahrgangsstufe mittels 43 Multiple-Choice-Items erhoben (Eigenentwicklungen, die teils in Kooperation mit dem Institut für Qualitätsentwicklung (Leseverständnistest 7 Hessen, 2007) entwickelt wurden). In der sechsten Jahrgangsstufe wurde ein 31 Multiple-Choice-Items umfassender Lesekompetenztest (Eigenentwicklung) eingesetzt. Die Schülerinnen und Schüler mussten die jeweiligen Texte lesen, relevante Informationen im Text suchen sowie entsprechende Inferenzen aus dem Text generieren, um die Multiple-Choice-Items zu beantworten. Die Lesekompetenztests waren über ein Anker-Item-Test-Design (common item design with nonequivalent groups/anchor-item test design; vgl. Holland, Dorans & Petersen, 2007; Kolen & Brennan, 2004) miteinander verbunden, welches die Schätzung von Personenfähigkeiten auf einer gemeinsamen Metrik ermöglicht. In einem ersten Schritt wurden hierfür die Schwierigkeitsparameter aller Items zu t₁ unter Zugrundelegung eines 1-parametrischen Raschmodells mit der Software ConQuest (vgl. Wu, Adams, Wilson & Haldane, 2007) geschätzt und für die sich anschließende Skalierung des zweiten Messzeitpunktes fixiert. Die Fähigkeiten der Personen wurden mit Hilfe von Weighted-Likelihood-Estimates (WLEs) geschätzt. Im darauf folgenden Schritt wurden die Schwierigkeitsparameter der Lesekompetenzitems zu t₂ unter Verwendung der fixierten Schwierigkeitsparameter der sich überlappenden Items aus t₁ geschätzt (separate estimation) und erneut WLEs für die Personenfähigkeiten zu t₂ bestimmt. Im Hinblick auf die kriterienbezogene Validität zeigte sich für den Lesekompetenztest zu t₁ eine Korrelation mit der Deutschnote aus dem Halbjahreszeugnis der fünften Klasse von $r = -.39$. Die Korrelation zwischen Lesekompetenztest zu t₂ und Deutschnote aus dem Halbjahreszeugnis der sechsten Klasse lag bei $r = -.40$. Die interne Konsistenz lag zu t₁ bei $\alpha = .79$ und zu t₂ bei $\alpha = .82$.

3.2.1.2 Mathematische Kompetenz. Die mathematische Kompetenz der Schülerinnen und Schüler wurde am Ende der fünften Jahrgangsstufe mittels 44 Items und am Ende der sechsten Jahrgangsstufe über einen 40 Items umfassenden Mathematiktest erfasst. Die Items stammten aus dem Deutschen Mathematiktest für fünfte und sechste Klassen (DEMAT 5+; vgl. Marx & Opitz-Karig, 2005; DEMAT 6+, vgl. Marx & Krockner, in Vorbereitung), dem Projekt zur Analyse der Leistungsentwicklung in Mathematik (PALMA; vgl. Pekrun et al., 2002), der dritten internationalen Mathematik- und Naturwissenschaftsstudie (TIMSS; vgl. Mullis et al., 1997) sowie aus Eigenentwicklung. Die beiden Tests umfassten dabei Aufgaben aus den Bereichen Arithmetik, Sachrechnen und Geometrie. Die Mathematiktests waren, ebenso wie die Verfahren zur Messung der Lesekompetenz, über ein Anker-Item-Test-Design miteinander verbunden. Die mathematischen

Fähigkeiten der Personen zu den beiden Messzeitpunkten wurden, ähnlich wie bereits für die Lesekompetenz geschildert, auf einer gemeinsamen Metrik durch Weighted-Likelihood-Estimates geschätzt. Eine Interpretation des mittleren längsschnittlichen Kompetenzzugewinns zwischen den beiden Messzeitpunkten ist jedoch unter anderem aufgrund der relativ hohen Curriculumsnähe der eingesetzten Mathematiktests kritisch zu sehen. Ferner erweisen sich die Kompetenzzuwächse in der Regel sensitiv gegenüber dem eingesetzten Skalierungsverfahren (vgl. Wu, 2010). Diese Einschränkung trifft auf die Ergebnisse des Lese- sowie Mathematiktests gleichermaßen zu. Die interne Konsistenz betrug zu t1 $\alpha = .87$ und zu t2 $\alpha = .89$.

3.2.1.3 Kognitive Grundfähigkeiten. Zur Erfassung der kognitiven Grundfähigkeiten der Schülerinnen und Schüler wurden am Ende der fünften Jahrgangsstufe 15 Items aus dem Subtest „Matrizen“ aus der Grundintelligenztest Skala 2 eingesetzt (CFT-20-R, vgl. Weiß, 2006). Dieser Test erfasst die Fähigkeit, Regeln und Zusammenhänge bei figuralen Problemstellungen zu erkennen. Die kognitiven Grundfähigkeiten wurden durch die Summe der richtigen Lösungen angegeben. Laut Testmanual weist dieser Test zufriedenstellende psychometrische Eigenschaften auf (Korrelation zwischen Subtest „Matrizen“ und Gesamtwert $r = .82$).

Die Mittelwerte und Standardabweichungen der Kompetenzmaße sind in Tabelle 1 für die vollständige Analysestichprobe sowie getrennt nach der Schulform dargestellt.

Tabelle 1: Mittelwerte und Standardabweichungen der Schülerleistungen in der Analysestichprobe und getrennt nach Schulform zu den zwei Messzeitpunkten

	Gesamt (N = 502)				Haupt- und Realschulen (N = 113)				Gymnasien (N = 389)			
	t1		t2		t1		t2		t1		t2	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Lesekompetenz	0.91	0.77	1.23	0.99	0.31	0.72	0.47	0.74	1.08	0.69	1.45	0.95
Mathematische Kompetenz	0.15	1.02	1.08	1.22	-0.65	1.01	0.06	0.99	0.38	0.86	1.32	1.11
Kognitive Fähigkeiten (t1)	11.45	2.10	–	–	10.64	2.18	–	–	11.76	1.97	–	–

Anmerkungen. Lese- und mathematische Kompetenz: WLE-Scores; kognitive Grundfähigkeiten: Rohwerte.

3.2.1.4 Sozioökonomischer Status in der Familie. Die Daten zum sozialen Hintergrund der Schülerinnen und Schüler wurden im Rahmen von standardisierten telefonischen Elterninterviews erhoben. Im Rahmen der Analysen wurde der höchste sozioökonomische Status in der Familie verwendet (HISEI; vgl. Ganzeboom, De Graaf & Treiman, 1992). Der mittlere HISEI zu t1 betrug 55.8 Punkte ($SD = 16.7$, $Min = 16$, $Max = 90$). Für die Schülerinnen und Schüler der Haupt- und Realschulen lag der mittlere HISEI bei $M = 44.1$ ($SD = 14.3$) und bei den Gymnasiastinnen und Gymnasiasten lag der mittlere HISEI bei $M = 58.5$.

($SD = 15.8$). Der Bundesdurchschnitt für den HISEI aus dem Mikrozensus 2008 (vgl. Nold, 2010) liegt dagegen bei $M = 47.6$ (Hauptschule: $M = 39.3$, Realschule $M = 44.8$, Gymnasium $M = 55.1$).

3.2.2 Lehrervariablen

3.2.2.1 Diagnostische Kompetenz der Deutsch- und Mathematiklehrkräfte. Zur Erfassung der diagnostischen Kompetenz wurde den Lehrkräften am Ende der fünften Jahrgangsstufe (t_1) ein Fragebogen vorgelegt. Der Fragebogen bestand aus einem einseitigen Prosatext mit sieben Multiple-Choice-Aufgaben aus dem Lesekompetenztest sowie aus sieben Items aus dem Mathematiktest, die alle an der Studie teilnehmenden Schülerinnen und Schüler bearbeitet hatten. Um den Bearbeitungsaufwand für die Lehrkräfte gering zu halten, wurden für die Lehrereinschätzungen sieben Schülerinnen und Schüler pro Klasse zufällig ausgewählt. Die Lehrkräfte wurden gebeten, für jede der sieben Aufgaben aus dem jeweiligen Kompetenztest anzugeben, ob die genannte Schülerin bzw. der genannte Schüler die jeweilige Aufgabe ihrer Einschätzung nach lösen konnte oder nicht. Im Hinblick auf die Validität zeigte sich für die Einschätzungen der Deutschlehrkräfte eine Korrelation mit der Deutschnote aus dem Halbjahreszeugnis der fünften Klasse von $r = -.52$. Die Korrelation der Einschätzungen der Mathematiklehrkräfte mit der Mathematiknote aus dem Halbjahreszeugnis der fünften Klasse betrug $r = -.60$. Anhand dieser aufgabenspezifischen Einschätzungen wurden die Rangordnungskomponente und der aufgabenspezifische Treffer berechnet. Für die Rangordnungskomponente wurde zunächst für jede Schülerin bzw. jeden Schüler das individuelle Lehrerurteil als Anzahl erwarteter Lösungen und die individuelle Schülerleistung als Anzahl tatsächlicher Lösungen ermittelt. Dann wurden klassenweise Korrelationskoeffizienten nach Pearson zwischen den Lehrereinschätzungen und den Schülerleistungen berechnet, die anschließend an eine Fisher Z-Transformation gemittelt wurden. Der Korrelationskoeffizient nach Pearson ist ein strengeres Kriterium als der Rangkorrelationskoeffizient nach Spearman, da neben der Fähigkeit der Lehrkraft, die reine Rangposition einzelner Schülerinnen bzw. Schüler innerhalb der Klasse zutreffend einzuschätzen auch die Fähigkeit, den relativen Leistungsabstand zu anderen Schülerinnen und Schülern korrekt einzuschätzen, eingehen (vgl. Schrader, 1989). Zur Berechnung des aufgabenspezifischen Treffers wurde für jede der sieben Aufgaben des jeweiligen Kompetenztests ermittelt, ob die von der Lehrkraft als gelöst (bzw. nicht gelöst) angegebene Aufgabe auch tatsächlich dieselbe Aufgabe ist, die die Schülerin bzw. der Schüler löste (bzw. nicht löste). Bei korrekter Einschätzung gab es jeweils einen Trefferpunkt. Der aufgabenspezifische Treffer wurde dann als Quotient aus dem Summenwert der Treffer über alle Schüler und der Anzahl der Aufgaben berechnet. Nimmt die Trefferquote den Wert 0 an, gab die Lehrkraft in keinem der Fälle eine

korrekte Einschätzung ab. Eine Trefferquote von 1 zeigt an, dass die Lehrkraft alle Schülerinnen und Schüler bei allen Aufgaben korrekt einschätzte.¹

3.2.2.2 Individualisierung des Unterrichts. Die Individualisierung im Deutsch- bzw. Mathematikunterricht durch die Lehrkräfte wurde zu t1 anhand von je vier Items erhoben (adaptiert nach Clausen, 2002, z.B.: „Ich gebe Kindern je nach Leistung unterschiedlich schwere Aufgaben.“). Die Items waren auf einer vierstufigen Ratingskala von 1 = „trifft nicht zu“ bis 4 = „trifft zu“ zu beantworten. Für die Individualisierung im Deutschunterricht lag die interne Konsistenz bei $\alpha = .85$ und für den Mathematikunterricht bei $\alpha = .79$.

3.2.2.3 Strukturierungshilfen. Die Strukturierungshilfen der Lehrkräfte im Deutsch- bzw. Mathematikunterricht wurden zu t1 mit je drei Items anhand einer Ratingskala von 1 = „nie“ bis 4 = „oft“ erfasst (in Anlehnung an Rakoczy, Buff & Lipowsky, 2005 und Eigenentwicklung, z.B.: „Ich fasste noch einmal den Stoff zusammen, damit die Schüler dieser Klasse ihn sich gut merken konnten.“). Die Reliabilitäten waren zufriedenstellend (Deutsch: Cronbachs $\alpha = .80$, Mathematik: Cronbachs $\alpha = .79$).

Ein Überblick über die Mittelwerte und Standardabweichungen für die Individualisierung des Unterrichts und die Strukturierungshilfen der Deutsch- und Mathematiklehrkräfte wird in Tabelle 2 sowohl für die vollständige Analysestichprobe als auch getrennt nach Schulform gegeben.

Tabelle 2: Mittelwerte und Standardabweichungen der Individualisierung des Unterrichts und der Strukturierungshilfen der Deutsch- und Mathematiklehrkräfte in der Analysestichprobe sowie getrennt nach Schulform

	Gesamt		Haupt- und Realschulen		Gymnasien	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Deutschlehrkräfte						
Individualisierung	2.28	0.74	2.84	0.82	2.07	0.59
Strukturierungshilfen	3.43	0.71	3.79	0.40	3.30	0.76
Mathematiklehrkräfte						
Individualisierung	3.47	0.70	3.36	0.32	3.81	0.18
Strukturierungshilfen	2.71	0.65	2.51	0.60	3.36	0.77

Anmerkungen. Deutschlehrkräfte: Gesamt: $N = 40$, Haupt- und Realschulen: $N = 11$, Gymnasien: $N = 29$; Mathematiklehrkräfte: Gesamt: $N = 29$, Haupt- und Realschulen: $N = 7$, Gymnasien: $N = 22$.

¹ Die Bildung der beiden aufgabenspezifischen Indikatoren wird ausführlich anhand eines Beispiels bei Karing et al. (2011) beschrieben.

3.3 Statistische Analysen

3.3.1 Umgang mit fehlenden Werten

Zur optimalen Ausschöpfung unserer Stichprobe wurden die fehlenden Werte bei den Schülerinnen und Schülern im Datensatz mittels Multipler Imputation geschätzt ($m = 5$; vgl. Lüdtke, Robitzsch, Trautwein & Köller, 2007). Dafür wurde die Software SPSS (IBM SPSS Statistics 19) verwendet.

3.3.2 Mehrebenenanalysen

Für die Überprüfung der zweiten und dritten Fragestellung ist die hierarchische Datenstruktur zu berücksichtigen. Ein adäquates Auswertungsverfahren stellt die Mehrebenenanalyse dar, die auch eine simultane Einbeziehung der Prädiktoren auf Schüler- und Klassenebene ermöglicht (vgl. Raudenbush & Bryk, 2002). Die Mehrebenenanalysen wurden für jeden der fünf Datensätze mittels des Programms HLM 6.08 durchgeführt (vgl. Raudenbush, Bryk, Cheong & Congdon, 2004). Alle dargestellten Modelle sind sogenannte *random intercept fixed slope* Modelle. Die Schätzungen erfolgten mit der restricted maximum likelihood Methode (REML). Um inhaltlich interpretierbare Werte zu erhalten, wurden die metrischen Prädiktoren vor der Schätzung der Modelle z-standardisiert und die dichotomen Variablen wurden dummy-kodiert (Kodierung mit 0 und 1). Die Kodierung der Schulform erfolgte anhand einer Dummy-Variable (0 = Hauptschulen und Realschulen, 1 = Gymnasien). Die Haupt- und Realschulklassen wurden aufgrund der geringen Anzahl in einer Kategorie zusammengefasst.

Die abhängige Variable war in allen Modellen die Schülerleistung (Lesekompetenz oder mathematische Kompetenz) in der sechsten Jahrgangsstufe (t2). Zunächst wurde ein Nullmodell berechnet, um festzustellen, ob überhaupt Varianz zwischen den Klassen bezüglich der Lesekompetenz bzw. der mathematischen Kompetenz vorhanden ist. Im nächsten Schritt wurde ein Modell spezifiziert, in dem die Kontrollvariablen auf Klassenebene (Schulform) und Schülerebene (Lesekompetenz bzw. mathematische Kompetenz zu t1, kognitive Grundfähigkeiten zu t1, HISEI zu t1, Geschlecht der Schülerinnen und Schüler) als Prädiktoren aufgenommen wurden (Kontrollmodell). Der Einfluss dieser Variablen wurde in den berechneten Mehrebenenmodellen kontrolliert, da überprüft werden sollte, ob die beiden Urteilskomponenten (aufgabenspezifischer Treffer, Rangordnungskomponente) über den Einfluss dieser Prädiktorvariablen hinaus mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz zusammenhängen. Die genannten Prädiktorvariablen wurden ausgewählt, weil diese mit der Lesekompetenz und der mathematischen Kompetenz zusammenhängen. So belegen internationale Schulleistungsstudien wie PISA und IGLU einerseits eine Überlegenheit der Mädchen im Bereich Lesen, andererseits weisen Mädchen im mathematischen Bereich schwächere Testleistungen auf als Jungen (z.B. Drechsel

& Artelt, 2007; Hornberg, Valtin, Potthoff, Schwippert & Schulz-Zander, 2007; Zimmer, Stick, Burba & Prenzel, 2006). Zudem zeigt sich in diesen Studien auch ein positiver Zusammenhang zwischen sozialer Herkunft und der Lesekompetenz bzw. mathematischen Kompetenz der Kinder (vgl. Baumert & Schümer, 2001; Bos, Schwippert & Stubbe, 2007). Darüber hinaus konnte in verschiedenen Studien ein positiver Einfluss der kognitiven Grundfähigkeiten auf die Lesekompetenz und die mathematische Kompetenz der Kinder gezeigt werden (vgl. Retelsdorf & Möller, 2008; Schaffner, Schiefele & Schneider, 2004; Stern & Möller, 2004).

Im nächsten Schritt wurden zusätzlich die beiden Indikatoren der diagnostischen Kompetenz separat auf Klassenebene ins Modell aufgenommen, um den Zusammenhang zwischen der diagnostischen Kompetenz der Lehrkräfte und der Leistungsentwicklung der Schülerinnen und Schüler zu überprüfen. Zur Überprüfung der Forschungsfrage, ob der Zusammenhang der diagnostischen Kompetenz der Lehrkräfte mit der Leistungsentwicklung durch die beiden Unterrichtsvariablen moderiert wird, wurde für beide Bereiche jeweils ein Mediansplit für die Individualisierung (Lesekompetenz: $Md = 2.25$; mathematische Kompetenz: $Md = 2.75$) und die Strukturierungshilfen (Lesekompetenz: $Md = 3.67$; mathematische Kompetenz: $Md = 3.67$) durchgeführt. Danach wurden für die vier Gruppen (niedrige und hohe Individualisierung, wenige und viele Strukturierungshilfen), jeweils für die Lesekompetenz und die mathematische Kompetenz, der aufgabenspezifische Treffer sowie die Rangordnungskomponente einzeln auf Klassenebene unter Konstanzhaltung der Kontrollvariablen in die Modelle aufgenommen.

4. Ergebnisse

4.1 Wie akkurat schätzen die Deutsch- und Mathematiklehrkräfte die Schülerleistungen ein?

In Tabelle 3 sind die beiden Indikatoren der diagnostischen Kompetenz dargestellt. Für die Deutschlehrkräfte zeigt sich eine durchschnittliche Trefferquote von $M = 0.66$ ($SD = 0.11$), während die mittlere Trefferquote bei den Mathematiklehrkräften bei $M = 0.58$ ($SD = 0.06$) liegt, dieser Unterschied erweist sich als signifikant ($t = 3.66$, $p < .05$, $d = 0.82$). Die Deutschlehrkräfte schätzen demzufolge im Durchschnitt 66 % der Schülerantworten im Bereich Lesekompetenz und die Mathematiklehrkräfte durchschnittlich 58 % der Schülerantworten in Mathematik korrekt ein. Für die Rangordnungskomponente auf Aufgabenebene lässt sich feststellen, dass die Deutschlehrkräfte lediglich eine durchschnittliche Korrelation von $\bar{r} = .19$ ($SD = 0.51$) aufweisen, während sich für die Mathematiklehrkräfte eine durchschnittliche Korrelation von $\bar{r} = .44$ ($SD = 0.38$) ergibt, wobei dieser Unterschied statistisch bedeutsam ist ($t = 2.29$, $p < .05$, $d = 0.56$). Wie aber an den Standardabweichungen in Tabelle 3 abzulesen ist, zeigt sich eine gro-

ße Variabilität zwischen den Lehrkräften in Bezug auf die Trefferquote und die Rangordnungskomponente.

Tabelle 3: Rangordnungskomponente und Aufgabenspezifischer Treffer für Deutsch- und Mathematiklehrkräfte

	<i>M</i>	<i>SD</i>	Min	Max
Deutschlehrkräfte				
Aufgabenspezifischer Treffer	0.66	0.11	0.40	0.86
Rangordnungskomponente	.19	0.51	-.85	.93
Mathematiklehrkräfte				
Aufgabenspezifischer Treffer	0.58	0.06	0.45	0.69
Rangordnungskomponente	.44	0.38	-.63	.79

Anmerkungen. *N* = 38–40 Deutschlehrkräfte, *N* = 29 Mathematiklehrkräfte.

4.2 Hängen die beiden Komponenten der diagnostischen Kompetenz mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz der Schülerinnen und Schüler zusammen?

Die Ergebnisse der Mehrebenenanalysen sind in Tabelle 4 und 5 wiedergegeben. Zunächst werden die Ergebnisse für die Lesekompetenz dargestellt (vgl. Tabelle 4). Die Lesekompetenz in der sechsten Jahrgangsstufe variiert zu 33.7 % zwischen den Klassen (Nullmodell). Die zur Kontrolle eingeführten Variablen auf Schüler- und Klassenebene erklären 40.2 % der Gesamtvarianz (Modell 1). Es zeigt sich, dass die Lesekompetenz am Ende der fünften Jahrgangsstufe der stärkste Prädiktor ist ($B = 0.37$, $p < .01$), aber auch das Geschlecht ($B = 0.29$, $p < .01$), die kognitiven Fähigkeiten ($B = 0.13$, $p < .01$) und der sozioökonomische Status in der Familie ($B = 0.11$, $p < .05$) hängen positiv und signifikant mit der Lesekompetenz am Ende der sechsten Jahrgangsstufe zusammen. Außerdem erweist sich die Schulform als statistisch bedeutsam ($B = 0.37$, $p < .01$). Im nächsten Schritt werden die beiden Indikatoren der diagnostischen Kompetenz einzeln auf Klassenebene unter Konstanthaltung der Kontrollvariablen in die Modelle aufgenommen. Für den aufgabenspezifischen Treffer zeigt sich ein signifikanter positiver Zusammenhang mit der Lesekompetenzentwicklung der Schülerinnen und Schüler ($B = 0.15$, $p < .05$, vgl. Modell 2),² während sich für die Rangordnungskomponente kein signifikan-

2 Die Analysen wurden nochmals nur für die Schulform Gymnasium durchgeführt, da die Gymnasiallehrkräfte eine signifikant höhere Trefferquote im Bereich Lesekompetenz aufweisen als die Haupt- und Realschullehrkräfte ($t = 3.92$, $p < .01$; Gymnasium: $M = 0.70$, $SD = 0.10$; Haupt-/Realschulen: $M = 0.56$, $SD = 0.09$). Es zeigt sich weiterhin ein in vergleichbarer Höhe, allerdings nur noch in der Tendenz, signifikanter positiver Zusammenhang des aufgabenspezifischen Treffers mit der Lesekompetenzentwicklung ($B = 0.17$, $p < .10$).

ter positiver Zusammenhang ergibt ($B = -0.02$, $p > .05$, vgl. Modell 3). Im Modell 2 werden 41.9 % und im Modell 3 40.7 % der Gesamtvarianz erklärt.

Tabelle 4: Ergebnisse der Mehrebenenanalysen zur Vorhersage der Lesekompetenz in der 6. Klassenstufe

	Null- modell	Modell 1		Modell 2		Modell 3 ^b	
		<i>B</i>	<i>SE (B)</i>	<i>B</i>	<i>SE (B)</i>	<i>B</i>	<i>SE (B)</i>
Klassenebene							
Schulform Gymnasium ^a		0.37**	0.11	0.19	0.18	0.37**	0.11
Aufgabenspezif. Treffer (t1)		–	–	0.15*	0.09	–	–
Rangordnungskomponente (t1)				–	–	-0.02	0.05
Schülerebene							
Lesekompetenz (t1)		0.37**	0.04	0.36**	0.04	0.36**	0.04
Kogn. Fähigkeiten (t1)		0.13**	0.03	0.13**	0.03	0.13**	0.03
HISEI (t1)		0.11*	0.05	0.10*	0.05	0.10*	0.05
Geschlecht		0.29**	0.07	0.29**	0.06	0.27**	0.06
Intercept (γ_{00})		0.77**	0.08	0.90**	0.11	0.77**	0.09
Residualvarianz							
Klassenebene (u_o)	0.337	0.081		0.065		0.090	
Schülerebene (r)	0.663	0.517		0.516		0.467	
ICC	0.337	0.135		0.112		0.162	
R^2		0.402		0.419		0.407	

Anmerkungen. ^a Referenz: Haupt- und Realschule.

Geschlecht: 0 = männlich, 1 = weiblich; 5 Hauptschulklassen und 6 Realschulklassen sowie 29 Gymnasialklassen, 502 Schülerinnen und Schüler. ^b Modell 3: 5 Hauptschulklassen und 6 Realschulklassen sowie 27 Gymnasialklassen, 476 Schülerinnen und Schüler.

ICC: Intraklassenkorrelationen (Varianz zwischen Klassen [u_0]/ Gesamtvarianz [$r + u_0$]).

* $p < .05$, ** $p < .01$.

Für die mathematische Kompetenz lässt sich feststellen, dass 47.7 % der Varianz der Mathematikleistung in der sechsten Jahrgangsstufe auf Unterschiede zwischen den Klassen zurückzuführen sind (Nullmodell, vgl. Tabelle 5). Wie Tabelle 5 (Modell 1) zeigt, erweisen sich nur die Mathematikleistung am Ende der fünften Jahrgangsstufe ($B = 0.81$, $p < .01$) und die Schulform ($B = 0.58$, $p < .01$) als statistisch bedeutsam. Die zur Kontrolle eingeführten Variablen auf Schüler- und Klassenebene erklären 63.0 % der Gesamtvarianz. Im nächsten Schritt werden der aufgabenspezifische Treffer (Modell 2) und die Rangordnungskomponente (Modell 3) einzeln auf Klassenebene unter Konstanthaltung der Kontrollvariablen in die Modelle aufgenommen. Es zeigt sich weder für den aufgabenspezifischen Treffer noch für die Rangordnungskomponente ein signifikanter positiver Zusammenhang mit der Leistungsentwicklung der Schülerinnen und Schüler in Mathematik (aufgabenspezifischer Treffer: $B = -0.02$, $p > .05$; Rangordnungskomponente: $B = -0.10$, $p > .05$). Insgesamt erklären im Modell 2 die berücksichtigten Prädiktoren 62.8 % und im Modell 3 63.3 % der Gesamtvarianz auf.

Tabelle 5: Ergebnisse der Mehrebenenanalysen zur Vorhersage der mathematischen Kompetenz in der 6. Klassenstufe

	Null- modell	Modell 1		Modell 2		Modell 3	
		<i>B</i>	<i>SE (B)</i>	<i>B</i>	<i>SE (B)</i>	<i>B</i>	<i>SE (B)</i>
Klassenebene							
Schulform Gymnasium ^a		0.58**	0.17	0.59**	0.16	0.58**	0.16
Aufgabenspezifischer Treffer (t1)		–	–	-0.02	0.08	–	–
Rangordnungskomponente (t1)				–	–	-0.10	0.05
Schülerebene							
Mathematische Kompetenz (t1)		0.81**	0.06	0.81**	0.06	0.81**	0.06
Kogn. Fähigkeiten (t1)		0.06	0.05	0.06	0.05	0.06	0.05
HISEI (t1)		0.03	0.05	0.03	0.05	0.02	0.05
Geschlecht		-0.01	0.10	-0.01	0.10	-0.01	0.10
Intercept (γ_{00})		0.64**	0.15	0.63**	0.16	0.64**	0.15
Residualvarianz							
Klassenebene (u_o)	0.881	0.098		0.102		0.093	
Schülerebene (r)	0.967	0.585		0.585		0.585	
ICC	0.477	0.143		0.148		0.137	
R^2		0.630		0.628		0.633	

Anmerkungen. ^a Referenz: Haupt- und Realschule; Geschlecht: 0 = männlich, 1 = weiblich; 4 Hauptschulklassen und 3 Realschulklassen sowie 22 Gymnasialklassen, 368 Schülerinnen und Schüler; ICC: Intraklassenkorrelationen (Varianz zwischen Klassen [u_o]/Gesamtvarianz [$r + u_o$]).

* $p < .05$, ** $p < .01$.

4.3 Werden die Zusammenhänge dieser beiden Urteils-komponenten mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz der Schülerinnen und Schüler durch die Unterrichtsvariablen Individualisierung und Strukturierungshilfen moderiert?

Zur Beantwortung dieser Fragestellung werden Mehrebenenanalysen getrennt nach hoher und niedriger Individualisierung sowie nach der Gabe von wenigen bzw. vielen Strukturierungshilfen durchgeführt. Für die vier Gruppen wird der Zusammenhang der beiden Urteils-komponenten mit der Leistungsentwicklung überprüft, indem diese beiden Indikatoren einzeln auf Klassenebene unter Konstanthaltung der Kontrollvariablen auf Schüler- und Klassenebene in die Mehrebenenmodelle aufgenommen werden. Zunächst werden die Befunde für die Lesekompetenz dargestellt. Die Koeffizienten dieser Modelle sind in den Tabellen 6 und 7 wiedergegeben. Für die Gruppe mit hoher Individualisierung ergibt sich ein signifikanter positiver Zusammenhang des aufgabenspezifischen Treffers mit der Lesekompetenzentwicklung ($B = 0.23$, $p < .05$), während in der Gruppe mit geringer Individualisierung der aufgabenspezifische Treffer nicht signifikant mit der Lesekompetenzentwicklung ($B = 0.10$, $p > .05$) zusammenhängt. Dagegen hängt die

Rangordnungskomponente weder in der Gruppe mit hoher Individualisierung noch in der Gruppe mit niedriger Individualisierung signifikant mit der Entwicklung der Lesekompetenz zusammen (geringe Individualisierung: $B = -0.04$, $p > .05$; hohe Individualisierung: $B = 0.03$, $p > .05$).

Bei den Strukturierungshilfen zeigt sich dagegen für den aufgabenspezifischen Treffer ein anderes Bild. In der Gruppe mit wenigen Strukturierungshilfen konnte ein signifikanter positiver Zusammenhang des aufgabenspezifischen Treffers mit der Lesekompetenzentwicklung ($B = 0.36$, $p < .01$) nachgewiesen werden, während sich kein signifikanter Zusammenhang in der Gruppe mit vielen Strukturierungshilfen ($B = 0.01$, $p > .05$) zeigt.³ Dagegen ergibt sich für die Rangordnungskomponente kein signifikanter Zusammenhang in beiden Gruppen (wenige Strukturierungshilfen: $B = 0.05$, $p > .05$; viele Strukturierungshilfen: $B = -0.07$, $p > .05$).

3 Die Analysen wurden nochmals nur für die Schulform Gymnasium durchgeführt, weil sich für die Gymnasiallehrkräfte eine signifikant höhere Trefferquote im Bereich Lesekompetenz zeigt als für die Haupt- und Realschullehrkräfte. Der signifikante positive Zusammenhang des aufgabenspezifischen Treffers mit der Lesekompetenzentwicklung bleibt in den Gruppen hohe Individualisierung ($B = 0.35$, $p < .01$) und wenige Strukturierungshilfen ($B = 0.36$, $p < .01$) erhalten. Für die anderen beiden Gruppen zeigen sich hier keine signifikanten Zusammenhänge (ohne Tabelle).

Tabelle 6: Ergebnisse der Mehrebenenanalysen zur Vorhersage der Lesekompetenz in der 6. Klassenstufe getrennt nach geringer und hoher Individualisierung

	geringe Individualisierung				hohe Individualisierung			
	Modell 1		Modell 2 ^b		Modell 1		Modell 2	
	<i>B</i>	<i>SE (B)</i>	<i>B</i>	<i>SE (B)</i>	<i>B</i>	<i>SE (B)</i>	<i>B</i>	<i>SE (B)</i>
Klassenebene								
Schulform Gymnasium ^a	0.39	0.19	0.50**	0.16	-0.02	0.21	0.31*	0.13
Aufgabenspezifischer Treffer (t1)	0.10	0.09	–	–	0.23*	0.09	–	–
Rangordnungs- komponente (t1)	–	–	-0.04	0.04	–	–	0.03	0.11
Schülerebene								
Lesekompetenz (t1)	0.42**	0.08	0.41**	0.08	0.30**	0.04	0.32**	0.04
Kogn. Fähigkeiten (t1)	0.16**	0.05	0.19**	0.05	0.11**	0.04	0.10**	0.04
HISEI (t1)	0.00	0.09	-0.03	0.08	0.17**	0.04	0.18**	0.05
Geschlecht	0.28*	0.11	0.23	0.12	0.28**	0.07	0.30**	0.07
Intercept (γ_{00})	0.87**	0.10	0.81**	0.13	0.96**	0.16	0.72**	0.11
Residualvarianz								
Klassenebene (u_o)	0.054		0.083		0.047		0.081	
Schülerebene (r)	0.612		0.503		0.426		0.427	
ICC	0.081		0.142		0.099		0.159	
R^2	0.361		0.379		0.467		0.427	

Anmerkungen. ^a Referenz: Haupt- und Realschule.

Geschlecht: 0 = männlich, 1 = weiblich; hohe Individualisierung: 5 Hauptschulklassen und 2 Realschulklassen, 15 Gymnasialklassen, 267 Schülerinnen und Schüler; geringe Individualisierung Modell 1: 4 Realschulklassen und 14 Gymnasialklassen, 235 Schülerinnen und Schüler.

^b geringe Individualisierung Modell 2: 4 Realschulklassen und 12 Gymnasialklassen, 209 Schülerinnen und Schüler. ICC: Intraklassenkorrelationen (Varianz zwischen Klassen [u_o]/ Gesamtvarianz [$r + u_o$]).

* $p < .05$, ** $p < .01$.

Tabelle 7: Ergebnisse der Mehrebenenanalysen zur Vorhersage der Lesekompetenz in der 6. Klassenstufe getrennt nach wenigen und vielen Strukturierungshilfen

	wenige Strukturierungshilfen				viele Strukturierungshilfen			
	Modell 1		Modell 2 ^b		Modell 1		Modell 2 ^c	
	<i>B</i>	<i>SE (B)</i>	<i>B</i>	<i>SE (B)</i>	<i>B</i>	<i>SE (B)</i>	<i>B</i>	<i>SE (B)</i>
Klassenebene								
Schulform Gymnasium ^a	-0.14	0.22	0.33	0.19	0.36*	0.17	0.38*	0.15
Aufgabenspezifischer Treffer (t1)	0.36**	0.09	–	–	0.01	0.07	–	–
Rangordnungs- komponente (t1)	–	–	0.05	0.08	–	–	-0.07	0.05
Schülerebene								
Lesekompetenz (t1)	0.36**	0.08	0.34**	0.08	0.36**	0.05	0.36**	0.05
Kogn. Fähigkeiten (t1)	0.11*	0.05	0.12*	0.05	0.15**	0.03	0.14**	0.04
HISEI (t1)	0.06	0.06	0.07	0.06	0.15*	0.06	0.12	0.07
Geschlecht	0.20*	0.08	0.21	0.08	0.35**	0.09	0.31**	0.09
Intercept (γ_{00})	1.14**	0.19	0.80**	0.17	0.78**	0.09	0.77**	0.10
Residualvarianz								
Klassenebene (u_o)	0.046		0.165		0.041		0.047	
Schülerebene (r)	0.559		0.502		0.489		0.446	
ICC	0.076		0.248		0.077		0.095	
R^2	0.381		0.261		0.476		0.491	

Anmerkungen. ^a Referenz: Haupt- und Realschule.

Geschlecht: 0 = männlich, 1 = weiblich. wenige Strukturierungshilfen Modell 1: 2 Realschulklassen und 15 Gymnasialklassen, 213 Schülerinnen und Schüler.

^b wenige Strukturierungshilfen Modell 2: 2 Realschulklassen und 14 Gymnasialklassen, 204 Schülerinnen und Schüler. viele Strukturierungshilfen Modell 1: 5 Hauptschulklassen, 4 Realschulklassen und 14 Gymnasialklassen, 289 Schülerinnen und Schüler.

^c viele Strukturierungshilfen Modell 2: 5 Hauptschulklassen, 4 Realschulklassen und 13 Gymnasialklassen, 272 Schülerinnen und Schüler.

ICC: Intraklassenkorrelationen (Varianz zwischen Klassen [u_o]/ Gesamtvarianz [$r + u_o$]).

* $p < .05$, ** $p < .01$.

In den Tabellen 8 und 9 sind die Ergebnisse der Mehrebenenanalysen für den mathematischen Bereich dargestellt. Für den aufgabenspezifischen Treffer und die Rangordnungskomponente zeigt sich weder in der Gruppe mit geringer Individualisierung (aufgabenspezifischer Treffer: $B = -0.10$, $p > .05$, Rangordnungskomponente: $B = -0.11$, $p > .05$) noch in der Gruppe mit hoher Individualisierung (aufgabenspezifischer Treffer: $B = 0.11$, $p > .05$, Rangordnungskomponente: $B = -0.02$, $p > .05$) ein signifikanter Zusammenhang mit der Leistungsentwicklung der Schülerinnen und Schüler in Mathematik. Ein ähnliches Bild ergibt sich für die Strukturierungshilfen. Auch hier zeigt sich für beide Indikatoren der diagnostischen Kompetenz weder in der Gruppe mit wenigen Strukturierungshilfen (aufgabenspezifischer Treffer: $B = -0.11$, $p > .05$, Rangordnungskomponente: $B = -0.14$, $p > .05$) noch in der Gruppe mit vielen Strukturierungshilfen (aufgabenspezifischer Treffer: $B = 0.05$, $p > .05$, Rangordnungskomponente: $B = -0.03$, $p > .05$) ein signifikanter Zusammenhang mit der Leistungsentwicklung in Mathematik.

Tabelle 8: Ergebnisse der Mehrebenenanalysen zur Vorhersage der mathematischen Kompetenz in der 6. Klassenstufe getrennt nach geringer und hoher Individualisierung

	geringe Individualisierung				hohe Individualisierung			
	Modell 1		Modell 2		Modell 1		Modell 2	
	<i>B</i>	<i>SE (B)</i>	<i>B</i>	<i>SE (B)</i>	<i>B</i>	<i>SE (B)</i>	<i>B</i>	<i>SE (B)</i>
Klassenebene								
Schulform Gymnasium ^a	–	–	–	–	0.63**	0.19	0.69**	0.20
Aufgabenspezifischer Treffer (t1)	-0.10	0.08	–	–	0.11	0.12	–	–
Rangordnungs- komponente (t1)	–	–	-0.11	0.05	–	–	-0.02	0.21
Schülerebene								
Mathematische Kompetenz (t1)	0.80**	0.08	0.81**	0.08	0.81**	0.09	0.82**	0.09
Kogn. Fähigkeiten (t1)	0.11*	0.05	0.12*	0.05	-0.01	0.08	-0.01	0.08
HISEI (t1)	0.10	0.06	0.10	0.06	-0.07	0.06	-0.06	0.06
Geschlecht	0.06	0.14	0.06	0.13	-0.05	0.14	-0.05	0.14
Intercept (γ_{00})	1.20**	0.14	1.18**	0.12	0.59**	0.18	0.55**	0.16
Residualvarianz								
Klassenebene (u_o)	0.028		0.017		0.203		0.213	
Schülerebene (r)	0.547		0.546		0.619		0.619	
ICC	0.049		0.030		0.247		0.256	
R^2	0.425		0.437		0.612		0.608	

Anmerkungen. ^a Referenz: Haupt- und Realschule; Geschlecht: 0 = männlich, 1 = weiblich; geringe Individualisierung: 14 Gymnasialklassen, 196 Schülerinnen und Schüler; hohe Individualisierung: 4 Hauptschulklassen und 3 Realschulklassen, 8 Gymnasialklassen, 172 Schülerinnen und Schüler; ICC: Intraklassenkorrelationen (Varianz zwischen Klassen [u_o]/Gesamtvarianz [$r + u_o$]).

* $p < .05$, ** $p < .01$.

Tabelle 9: Ergebnisse der Mehrebenenanalysen zur Vorhersage der mathematischen Kompetenz in der 6. Klassenstufe getrennt nach wenigen und vielen Strukturierungshilfen

	wenige Strukturierungshilfen				viele Strukturierungshilfen			
	Modell 1		Modell 2		Modell 1		Modell 2	
	<i>B</i>	<i>SE (B)</i>	<i>B</i>	<i>SE (B)</i>	<i>B</i>	<i>SE (B)</i>	<i>B</i>	<i>SE (B)</i>
Klassenebene								
Schulform Gymnasium ^a	–	–	–	–	0.70**	0.19	0.71**	0.18
Aufgabenspezifischer Treffer (t1)	-0.11	0.16	–	–	0.05	0.08	–	–
Rangordnungskom- ponente (t1)	–	–	-0.14	0.06	–	–	-0.03	0.12
Schülerebene								
Mathematische Kompetenz (t1)	0.78**	0.08	0.79**	0.07	0.82**	0.09	0.82**	0.09
Kogn. Fähigkeiten (t1)	0.10	0.06	0.11	0.06	0.02	0.06	0.02	0.06
HISEI (t1)	0.01	0.06	0.01	0.06	0.04	0.07	0.04	0.07
Geschlecht	0.02	0.19	0.03	0.18	-0.03	0.11	-0.03	0.11
Intercept (γ_{00})	1.12**	0.18	1.08**	0.17	0.68**	0.15	0.66**	0.14
Residualvarianz								
Klassenebene (u_o)	0.108		0.077		0.088		0.091	
Schülerebene (r)	0.445		0.445		0.695		0.695	
ICC	0.195		0.148		0.112		0.116	
R^2	0.506		0.534		0.663		0.661	

Anmerkungen. ^a Referenz: Haupt- und Realschule; Geschlecht: 0 = männlich, 1 = weiblich; wenige Strukturierungshilfen: 11 Gymnasialklassen, 156 Schülerinnen und Schüler; viele Strukturierungshilfen: 4 Hauptschulklassen und 3 Realschulklassen, 11 Gymnasialklassen, 212 Schülerinnen und Schüler; ICC: Intraklassenkorrelationen (Varianz zwischen Klassen [u_o]/Gesamtvarianz [$r + u_o$]).

* $p < .05$, ** $p < .01$.

5. Diskussion

Im Rahmen dieses Beitrags wurde zum einen die Urteilsgüte von Deutsch- und Mathematiklehrkräften der Sekundarstufe I anhand des aufgabenspezifischen Treffers und der Rangordnungskomponente untersucht. Zum anderen wurde überprüft, ob diese beiden Urteilkomponenten mit der Entwicklung der Lesekompetenz und der mathematischen Kompetenz der Schülerinnen und Schüler einhergehen.

In Bezug auf den aufgabenspezifischen Treffer ergab sich, wie erwartet, eine moderate Trefferquote für die Lehrkräfte. Jedoch zeigte sich für die Deutschlehrkräfte eine signifikant höhere Trefferquote als für die Mathematiklehrkräfte. So schätzten die Deutschlehrkräfte im Mittel 66 % der Schülerantworten im Bereich Lesekompetenz korrekt ein, während die Mathematiklehrkräfte lediglich 58 %

der Schülerantworten zutreffend einschätzten. Auch in der COACTIV-Studie (vgl. Brunner et al., 2011) ergab sich bei den Mathematiklehrkräften der Sekundarstufe lediglich eine mittlere Trefferquote (für insgesamt zwei Aufgaben) von 51 %. Damit sind die mittleren Trefferquoten unserer Studie sowie diejenige in der COACTIV-Studie niedriger als die Trefferquoten in Studien bei Grundschullehrkräften im englischsprachigen Raum (z.B. Coladarci, 1986; Demaray & Elliott, 1998). Für die Rangordnungskomponente fielen die durchschnittlichen Korrelationen vergleichsweise niedrig aus, wobei sich bei den Mathematiklehrkräften signifikant höhere Korrelationen ergaben als bei den Deutschlehrkräften. Eine mögliche Erklärung für die sehr niedrige Korrelation im Bereich Lesen könnte sein, dass die sieben Aufgaben des Lesetests nur wenig zwischen den Schülerinnen und Schülern differenzierten, was wiederum die Rangreihenbildung bei den Schülerinnen und Schülern erschwerte. Darüber hinaus lagen sowohl die Befunde unserer Studie als auch derjenige von Anders und Kollegen (Anders et al., 2010; Median von $r = .35$) deutlich unter den Werten, die in Studien im Grundschulbereich berichtet wurden (z.B. Demaray & Elliott, 1998; Feinberg & Shapiro, 2009). So berichteten Demaray und Elliott (1998) Korrelationen von $r = .82$ für den Bereich Lesen und $r = .66$ für Mathematik. Allerdings schätzten in dieser Studie die Grundschullehrkräfte alle Items des Lesekompetenz- und Mathematiksubtests ein, während in unserer Arbeit die Grundlage für die Rangordnungskomponente die sieben Einschätzungen waren. Darüber hinaus wurden in unserer Arbeit klassenspezifische Korrelationen berechnet, die anschließend an eine Fisher Z-Transformation gemittelt wurden, während in der Studie von Demaray und Elliott (1998) nur Korrelationen für die Gesamtstichprobe ohne Berücksichtigung der Klassenebene für den jeweiligen Bereich ermittelt wurden. Dies kann wiederum die Höhe des Zusammenhangs zwischen Lehrerurteil und tatsächlicher Schülerleistung beeinflussen. Ursächlich für die gefundenen niedrigeren Werte der beiden Urteils Komponenten in der Sekundarstufe im Vergleich zur Grundschule könnten jedoch auch Unterschiede zwischen den Schulstufen sein. Erste Befunde dazu zeigten sich in der Studie von Karing (2009), in der globale Lehrereinschätzungen (Rangordnungskomponente) betrachtet wurden. In dieser Studie schätzten Grundschullehrkräfte die Lesekompetenz sowie die mathematische Kompetenz der Schülerinnen und Schüler signifikant akkurater ein als Gymnasiallehrkräfte. Dieser Befund ist mit der Annahme vereinbar, dass die Unterschiede in der Urteilsgenauigkeit zwischen Lehrkräften der Grundschule und der Sekundarstufe mit strukturellen Bedingungen (Zusammensetzung der Schülerschaft, Ausbildung, Klassenlehrer- vs. Fachlehrerprinzip) zusammenhängen (vgl. Karing, 2009; Schrader, 2009), wobei mögliche Einschränkungen der Befunde aufgrund von Varianzrestriktionen noch überprüft werden müssen. Darüber hinaus ergaben sich, wie bei bisherigen Studien im Grundschul- und Sekundarstufenbereich, auch in unserer Arbeit für beide Indikatoren der diagnostischen Kompetenz große Unterschiede zwischen den einzelnen Lehrkräften.

Unsere Analysen zum Zusammenhang der beiden Urteils Komponenten mit der Leistungsentwicklung der Schülerinnen und Schüler ergaben ein heteroge-

nes Befundmuster. Hypothesenkonform stellten wir für die Deutschlehrkräfte fest, dass der aufgabenspezifische Treffer einen signifikanten positiven Zusammenhang mit der Entwicklung der Lesekompetenz der Schülerinnen und Schüler aufwies, auch wenn relevante Variablen kontrolliert wurden. Darüber hinaus stellten wir in unserer Arbeit wie erwartet fest, dass dieser Zusammenhang durch Unterrichtsmerkmale moderiert wurde. So ging eine hohe Trefferquote in Kombination mit einer hohen Individualisierung des Unterrichts mit einer verstärkten Lesekompetenzentwicklung auf Seiten der Schülerinnen und Schüler einher. Dagegen zeigte sich ein unerwartetes Bild bei der Interaktion zwischen dem aufgabenspezifischen Treffer und der Gabe von Strukturierungshilfen durch die Lehrkräfte. Hier war die Lesekompetenzentwicklung am stärksten ausgeprägt, wenn eine hohe Trefferquote mit der Gabe von wenigen Strukturierungshilfen im Unterricht einherging. Eine mögliche Erklärung für diesen unerwarteten Befund könnte sein, dass leistungsfähigere Schülerinnen und Schüler weniger auf die Strukturierungsbemühungen der Lehrkräfte angewiesen sind, sondern bei ihnen insbesondere Formen des offenen Unterrichts und des selbstgesteuerten Lernens effektiver sind. So zeigte sich in der Studie von Möller, Jonen, Hardy und Stern (2002), dass leistungsstarke Grundschülerinnen und Grundschüler nicht auf eine stärkere Strukturierung der Lehr-Lernumgebung im Sachunterricht angewiesen waren, während leistungsschwache Schülerinnen und Schüler von einer stärkeren Strukturierung hinsichtlich ihres Lernerfolgs profitierten. Dies könnte möglicherweise die Unterschiede zwischen den Befunden in der vorliegenden Arbeit und der Studie von Schrader (1989; vgl. auch Schrader & Helmke, 1987) erklären. Während unsere Stichprobe fast 80 % Gymnasiastinnen und Gymnasiasten umfasste, untersuchte Schrader (1989; vgl. auch Schrader & Helmke, 1987) Hauptschülerinnen und Hauptschüler. In dieser Untersuchung zeigte sich, dass der Lernerfolg der Hauptschülerinnen und Hauptschüler am höchsten war, wenn hohe Diagnosekompetenz mit hoher Strukturierung des Unterrichts einherging, während sich unter der Bedingung hohe Diagnosekompetenz und geringe Strukturierung der niedrigste Lernerfolg ergab.

Für die Rangordnungskomponente ergab sich dagegen in dieser Studie bei den Deutschlehrkräften kein signifikanter positiver Zusammenhang mit der Entwicklung der Lesekompetenz der Schülerinnen und Schüler. Zudem zeigte sich für die Lesekompetenzentwicklung auch keine signifikante Interaktion zwischen der Rangordnungskomponente und den beiden Unterrichtsmerkmalen. Eine mögliche Erklärung für die unterschiedlichen Ergebnisse hinsichtlich der beiden Urteilskomponenten könnte darin bestehen, dass im Mittel die sich aus den Lehrerurteilen ergebende Rangreihe kaum derjenigen entsprach, die sich aus den tatsächlichen Schülerleistungen ergab und sich aufgrund der geringen Ausprägung dieser Komponente keine Zusammenhänge mit der Lesekompetenzentwicklung der Schülerinnen und Schüler zeigten. Denn nach Schrader (1989) muss dafür sowohl bei den Unterrichtsmerkmalen als auch bei den Komponenten der diagnostischen Kompetenz eine bestimmte Mindestausprägung vorliegen. Möglicherweise sind in unserer Stichprobe nicht ausreichend genug Lehrkräfte vorhanden, die

eine Mindestausprägung bei dieser Komponente aufweisen, um signifikante Zusammenhänge bzw. Interaktionen zu erreichen.

Des Weiteren stellten wir für die Mathematiklehrkräfte, entgegen unserer Annahmen, weder für die Rangordnungskomponente noch für den aufgabenspezifischen Treffer einen signifikanten positiven Zusammenhang mit der Entwicklung der mathematischen Kompetenz der Schülerinnen und Schüler fest, wobei dies mit den Befunden von Schrader (1989; vgl. auch Schrader & Helmke, 1987) einhergeht. In dieser Studie ergab sich für die Hauptschullehrkräfte ebenfalls kein signifikanter Zusammenhang zwischen den beiden Facetten der diagnostischen Kompetenz (personenbezogene und aufgabenbezogene Rangordnungskomponente) und dem Lernerfolg der Schülerinnen und Schüler der fünften Jahrgangsstufe im Fach Mathematik. Darüber hinaus zeigte sich in unserer Studie für die Leistungsentwicklung im Fach Mathematik auch keine signifikante Interaktion zwischen der Rangordnungskomponente bzw. dem aufgabenspezifischen Treffer und den Strukturierungshilfen bzw. der Individualisierung des Unterrichts. Eine mögliche Erklärung für die unterschiedlichen Befunde bei den Deutsch- und Mathematiklehrkräften hinsichtlich des aufgabenspezifischen Treffers könnte ebenfalls darin bestehen, dass eine bestimmte Mindestausprägung der Komponente notwendig ist, um signifikante Zusammenhänge bzw. Interaktionen zu erreichen (vgl. Schrader, 1989). Darüber hinaus wäre aber auch eine mögliche Erklärung für das Ausbleiben einer signifikanten Interaktion, dass die Mathematiklehrkräfte unserer Studie ihre diagnostischen Kompetenzen nicht dazu verwendeten, um Unterrichtsmerkmale adaptiv einzusetzen (vgl. Schrader, 1989). Zur Klärung dieser Deutungsmuster der Befunde sind jedoch weitere Arbeiten notwendig.

5.1 Einschränkungen, Implikationen und Ausblick

In dieser Studie beschränkt sich aus erhebungsökonomischen Gründen die diagnostische Kompetenz der Deutschlehrkräfte auf die Lesekompetenz, andere Teilkompetenzen des Faches Deutsch konnten nicht berücksichtigt werden. Des Weiteren muss hinsichtlich der beiden Urteilskomponenten berücksichtigt werden, dass eine perfekte Übereinstimmung zwischen Lehrereinschätzung und Schülerleistungen nicht zu erwarten ist, da zum einen die eingesetzten Kompetenztests eine einmalige Testperformanz widerspiegeln, während Lehrerurteile im schulischen Alltag auf Schülerleistungen in verschiedenen schriftlichen und mündlichen Leistungssituationen über einen längeren Zeitraum basieren. Ein weiteres Problem betrifft die Reliabilität der beiden Komponenten der diagnostischen Kompetenz, die – wie in anderen Studien auch – nicht direkt bestimmbar ist (vgl. McElvany et al., 2009; Schrader, 1989). Schrader (2009) macht in diesem Zusammenhang darauf aufmerksam, dass die Reliabilität der Diagnosekennwerte von der Reliabilität des Urteils und des Kriteriums abhängt, aber ohne durch sie eindeutig determiniert zu sein. Darüber hinaus ist auch kritisch anzumerken, dass in dieser Arbeit die beiden Unterrichtsmerkmale über Selbstauskünfte der

Lehrkräfte erfasst wurden und sich somit nur auf die Perspektive der Lehrkräfte beschränken, die wiederum mit möglichen Einschätzungstendenzen verbunden sein können. Ein weiterer Kritikpunkt ist zudem die Aufspaltung der beiden Unterrichtsvariablen am Median zur Beantwortung der dritten Fragestellung (vgl. Richter, 2007; Schrader, 1989). Eine feinere Differenzierung der Stichprobe zur Untersuchung von Moderatoreffekten ist jedoch aufgrund der geringen Stichprobengröße nicht sinnvoll. Die geringe Stichprobengröße von Haupt- und Realschulen schränkt zudem die Aussagekraft der berichteten Befunde dieser Studie für diese beiden Schulformen ein. Darüber hinaus legen Maas und Hox (2005) für die mehrebenen-analytische Auswertung eine Stichprobengröße von mindestens $N = 50$ Ebene-2-Einheiten nahe, da ansonsten die Ebene-2-Standardfehler verzerrt geschätzt werden. In unserer Studie liegt die Anzahl der Ebene-2-Einheiten unter $N = 50$, was wiederum auch mit einer geringeren Testpower verbunden ist. Demzufolge ist die Existenz von Effekten der diagnostischen Kompetenz in der Population, insbesondere im Bereich Mathematik, aufgrund eines hohen Stichprobenfehlers nicht auszuschließen.

Insgesamt liefert die vorliegende Studie einen Beitrag zum Zusammenhang zwischen der diagnostischen Kompetenz von Lehrkräften und der Leistungsentwicklung der Schülerinnen und Schüler. Im Gegensatz zu den bisherigen Studien betrachtet diese Arbeit neben dem mathematischen Bereich auch den Bereich Lesen. Darüber hinaus wird auch untersucht, ob der Zusammenhang zwischen den beiden Urteilskomponenten und dem Lernerfolg durch Unterrichtsvariablen modelliert wird. In zukünftigen Studien sollten jedoch auch andere Bereiche bzw. Fächer betrachtet werden. Zudem fehlen bislang Untersuchungen, die die Auswirkungen der diagnostischen Kompetenz in der Grundschule untersuchen.

Die in dieser wie auch in anderen Studien (z.B. Anders et al., 2010; McElvany et al., 2009; Schrader, 1989) festgestellten Defizite in der Urteilsgenauigkeit der Lehrkräfte, machen vor dem Hintergrund der Bedeutung der diagnostischen Kompetenz für effektiven Unterricht einen Trainingsbedarf für Lehrkräfte deutlich. Einen möglichen Ansatz bieten dafür „Vergleichsarbeiten“ (VERA, Helmke et al., 2004). Lehrkräfte erhalten hier Rückmeldungen über ihre aufgabenbezogene Diagnosegenauigkeit in den Fächern Deutsch und Mathematik. Die Befunde unserer Studie legen zudem nahe, auch aufgabenspezifische Urteile mit zu berücksichtigen. Allerdings weisen Helmke und Kollegen (2004) darauf hin, dass diese Rückmeldungen an die Lehrkräfte lediglich den Ausgangspunkt für mögliche Bemühungen zur Förderung der Akkuratheit der Einschätzungen von Lehrkräften darstellen, denn das bloße Informieren über die Genauigkeit ihrer Urteile dürfte wahrscheinlich nicht ausreichend sein, um die Urteilsgenauigkeit zu verbessern.

Danksagung

Diese Veröffentlichung wurde ermöglicht durch eine Sachbeihilfe der Deutschen Forschungsgemeinschaft (Kennz.: AR 301/6-1 und AR 301/6-2) im Bamberger Forschungsprojekt BiKS „Bildungsprozesse, Kompetenzentwicklung und Formation von Selektionsentscheidungen im Vor- und Grundschulalter (FOR 543)“. Wir danken den an der Studie teilnehmenden Schülerinnen, Schülern und Lehrkräften für ihre Teilnahme und allen im Rahmen der Datenerhebungen eingesetzten Studierenden für ihre engagierte Mitarbeit.

Literatur

- Anders, Y., Kunter, M., Brunner, M., Krauss, S. & Baumert, J. (2010). Diagnostische Fähigkeiten von Mathematiklehrkräften und ihre Auswirkungen auf die Leistungen ihrer Schülerinnen und Schüler. *Psychologie in Erziehung und Unterricht*, 57, 175–193.
- Artelt, C. & Gräsel, C. (2009). Gasteditorial. Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23, 157–160.
- Artelt, C., Stanat, P., Schneider, W. & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69–137). Opladen: Leske + Budrich.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9, 469–520.
- Baumert, J. & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 323–407). Opladen: Leske + Budrich.
- Bos, W., Schwippert, K. & Stubbe, T. C. (2007). Die Koppelung von sozialer Herkunft und Schülerleistung im internationalen Vergleich. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006. Lesekompetenz von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 225–248). Münster: Waxmann.
- Brunner, M., Anders, Y., Hachfeld, A. & Krauss, S. (2011). Diagnostische Fähigkeiten von Mathematiklehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 215–234). Münster: Waxmann.
- Clausen, M. (2002). *Qualität von Unterricht – Eine Frage der Perspektive?* Münster: Waxmann.
- Coladarci, T. (1986). The accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology*, 78, 141–146.
- Demaray, M. K. & Elliott, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, 13, 8–24.
- Drechsel, B. & Artelt, C. (2007). Lesekompetenz. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Hrsg.), *PISA 2006. Die*

- Ergebnisse der dritten internationalen Vergleichsstudie* (S. 225–248). Münster: Waxmann.
- Feinberg, A. B. & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research*, 102, 453–462.
- Ganzeboom, H. B. G., De Graaf, P. M. & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21, 1–56.
- Haag, L. & Lohrmann, K. (2007). Diagnostische (In-)Kompetenz von Lehrern. In B. Fuchs & C. Schönherr (Hrsg.), *Urteilkraft und Pädagogik. Beiträge zu einer pädagogischen Handlungstheorie* (S. 239–249). Würzburg: Königshausen & Neumann.
- Helmke, A., Hosenfeld, I. & Schrader, F. W. (2003). Diagnosekompetenz in Ausbildung und Beruf entwickeln. *Karlsruher pädagogische Beiträge*, 55, 15–34.
- Helmke, A., Hosenfeld, I. & Schrader, F. W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Grieser (Hrsg.), *Schulleitung und Schulentwicklung* (S. 119–144). Hohengehren: Schneider Verlag.
- Hoge, R. D. & Coladarchi, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Psychology*, 59, 297–313.
- Holland, P. W., Dorans, N. J. & Petersen, N. S. (2007). Equating test scores. In C. R. Rao & S. Sinharay (Hrsg.), *Handbook of statistics* (Bd. 26: Psychometrics, S. 169–203). Amsterdam: Elsevier.
- Hornberg, S., Valtin, R., Potthoff, B., Schwippert, K. & Schulz-Zander, R. (2007). Lesekompetenzen von Mädchen und Jungen im internationalen Vergleich. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 195–223). Münster: Waxmann.
- Ingenkamp, K. (1992). *Lehrbuch der Pädagogischen Diagnostik*. Weinheim: Beltz.
- Institut für Qualitätsentwicklung (2007). *Leseverständnistest 7 Hessen*. Wiesbaden: Institut für Qualitätsentwicklung.
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23, 197–209.
- Karing, C., Matthäi, J. & Artelt, C. (2011). Genauigkeit von Lehrerurteilen über die Lesekompetenz ihrer Schülerinnen und Schüler in der Sekundarstufe I – Eine Frage der Spezifität? *Zeitschrift für Pädagogische Psychologie*, 25, 159–172.
- Kolen, M. J. & Brennan, R. L. (2004). *Test equating, scaling, and linking. Methods and practices*. New York: Springer.
- Kuntze, S. (2006). *Konzeption einer Lernumgebung für den gymnasialen Mathematikunterricht und Evaluation einer Themenstudienarbeit zum mathematischen Beweisen und Argumentieren*. Zugriff am 26.07.2011 unter http://edoc.ub.uni-muenchen.de/6278/1/Kuntze_Sebastian.pdf
- Lehmann, R. H., Peek, R., Gänsfuß, R., Lutkat, S., Mücke, S. & Barth, I. (2000). *Qualitätsuntersuchungen an Schulen zum Unterricht in Mathematik (QuaSUM)*. Potsdam: Ministerium für Bildung, Jugend und Sport des Landes Brandenburg.
- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. Probleme und Lösungen. *Psychologische Rundschau*, 58, 103–117.
- Maas, C. J. M. & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 1, 85–91.
- Marx, H. & Krockner, N. (in Vorbereitung). *Deutscher Mathematiktest für 6. Klassen*. Manuskript in Vorbereitung.

- Marx, H. & Opitz-Karig, U. (2005). *DEMAT5+: Deutscher Mathematiktest für fünfte Klassen*. Unveröffentlicht, Universität Leipzig.
- McElvany, N., Schroeder, S., Richter, T., Hachfeld, A., Baumert, J., Schnotz, W., Horz, H. & Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern. *Zeitschrift für Pädagogische Psychologie*, 23, 223–235.
- Möller, K., Jonen, A., Hardy, I. & Stern, E. (2002). Die Förderung von naturwissenschaftlichem Verständnis bei Grundschulkindern durch Strukturierung der Lernumgebung. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (S. 176–191). Weinheim: Beltz. (= Zeitschrift für Pädagogik. 45. Beiheft).
- Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L. & Smith, T. A. (1997). *Mathematics achievement in the primary school years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.
- Nold, D. (2010). *Sozioökonomischer Status von Schülerinnen und Schülern 2008. Ergebnisse des Mikrozensus*. Statistisches Bundesamt. Zugriff am 26.07.2011 unter http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Querschnittsveroeffentlichungen/WirtschaftStatistik/BildungForschungKultur/StatusSchueler__22010,property=file.pdf
- Pekrun, R., Götz, T., Jullien, S., Zirngibl, A., von Hofe, R. & Blum, W. (2002). *Skalenhandbuch PALMA 1. Messzeitpunkt (5. Klassenstufe)*. Universität München: Institut für Pädagogische Psychologie.
- Rakoczy, K., Buff, A. & Lipowsky, F. (2005). Teil 1: Befragungsinstrumente. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“*. Frankfurt a. M.: Gesellschaft zur Förderung Pädagogischer Forschung/Deutsches Institut für Internationale Pädagogische Forschung.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models* (2. Aufl.). Thousand Oaks, CA: Sage.
- Raudenbush, S. W., Bryk, A., Cheong, Y. F. & Congdon, R. (2004). *HLM 6: Hierarchical linear and nonlinear modeling*. Chicago, IL: Scientific Software International.
- Retelsdorf, J. & Möller, J. (2008). Entwicklungen von Lesekompetenz und Lesemotivation. Schereneffekte in der Sekundarstufe? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 40, 179–188.
- Richter, T. (2007). Wie analysiert man Interaktionen von metrischen und kategorialen Prädiktoren? Nicht mit Median-Splits! *Zeitschrift für Medienpsychologie*, 19, 116–125.
- Schaffner, E., Schiefele, U. & Schneider, W. (2004). Ein erweitertes Verständnis der Lesekompetenz: Die Ergebnisse des nationalen Ergänzungstests. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (S. 197–242). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Frankfurt a. M.: Peter Lang.
- Schrader, F.-W. (2009). Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23, 237–245.
- Schrader, F.-W. & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1, 27–52.

- Schrader, F.-W. & Helmke, A. (2001). Alltägliche Leistungsbeurteilung durch Lehrer. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 45–58). Weinheim: Beltz.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19, 85–95.
- Ständiges Sekretariat der Kultusminister der Länder der Bundesrepublik Deutschland (KMK). (2004). *Standards für die Lehrerbildung*. Zugriff am 26.07.2011 unter http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf
- Stern, E. & Möller, K. (2004). Der Erwerb anschlussfähigen Wissens als Ziel des Grundschulunterrichtes. In D. Lenzen, J. Baumert, R. Watermann & U. Trautwein (Hrsg.), *PISA und die Konsequenzen für die erziehungswissenschaftliche Forschung* (S. 25–36). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Südkamp, A., Kaiser, J. & Möller, J. (2011). *Accuracy of teachers' judgments of students' academic achievement: A meta-analysis*. Manuskript eingereicht zur Publikation.
- Südkamp, A., Möller, J. & Pohlmann, B. (2008). Der Simulierte Klassenraum. Eine experimentelle Untersuchung zur diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 22, 261–276.
- Weinert, F. E., Schrader, F.-W. & Helmke, A. (1990). Educational expertise: Closing the gap between educational research and classroom practice. *School Psychology International*, 11, 163–180.
- Weiß, R. H. (2006). *Grundintelligenztest Skala 2 – Revision (CFT 20-R)*. Göttingen: Hogrefe.
- Wu, M. L. (2010). Measurement, sampling, and equating errors in large-scale assessments. *Educational Measurement: Issues and Practice*, 29 (4), 15–27.
- Wu, M. L., Adams, R. J., Wilson, M. & Haldane, S. A. (2007). *ACER ConQuest version 2.0: Generalised item response modelling software*. Camberwell: ACER Press.
- Zimmer, K., Stick, A., Burba, D. & Prenzel, M. (2006). PISA 2003 – Kompetenzmuster von Jungen und Mädchen in den deutschen Ländern. *Unterrichtswissenschaft*, 34, 310–329.